

Docket No.: HI-0074

#2  
LTY  
03/18/02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Jae Shin YU, Sung Bae JUN and Kyoung Ro YOON :

Serial No.: New U.S. Patent Application

Filed: March 7, 2002

For: METHOD FOR GENERATING SYNTHETIC KEY FRAME BASED  
UPON VIDEO TEXT

JC979 U.S. PTO  
10/091472  
03/07/02

**TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT**

Assistant Commissioner of Patents  
Washington, D. C. 20231

Sir:

At the time the above application was filed, priority was claimed based on the  
following application:

Korean Patent Application No. 12184/2001, filed March 9, 2001.

A copy of each priority application listed above is enclosed.

Respectfully submitted,  
FLESHNER & KIM, LLP

*Carl R. Wesolowski*

Daniel Y.J. Kim  
Registration No. 36,186  
Carl R. Wesolowski  
Registration No. 40,372

P. O. Box 221200  
Chantilly, Virginia 20153-1200  
703 502-9440

**Date: March 7, 2002**

DYK/CRW/cng

#2

JC979 U.S. PTO  
10/091472  
03/07/02



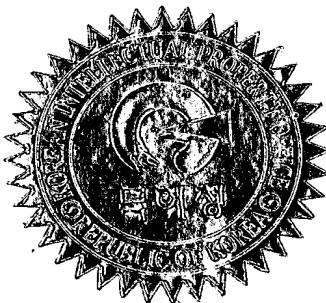
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto  
is a true copy from the records of the Korean Intellectual  
Property Office.

출원번호 : 특허출원 2001년 제 12184 호  
Application Number PATENT-2001-0012184

출원년월일 : 2001년 03월 09일  
Date of Application MAR 09, 2001

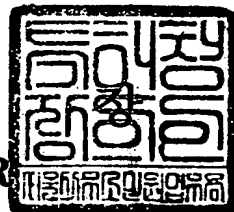
출원인 : 엘지전자주식회사  
Applicant(s) LG ELECTRONICS INC.



2001 년 12 월 12 일

특 허 청

COMMISSIONER



## 【서지사항】

【서류명】 특허출원서  
【권리구분】 특허  
【수신처】 특허청장  
【참조번호】 0001  
【제출일자】 2001.03.09  
【국제특허분류】 G06F  
【발명의 명칭】 비디오 텍스트 합성 키 프레임 추출방법  
【발명의 영문명칭】 METHOD FOR DETECTING CAPTION SYNTHETIC KEY FRAME IN VIDEO STREAM

## 【출원인】

【명칭】 엘지전자 주식회사  
【출원인코드】 1-1998-000275-8

## 【대리인】

【성명】 허용록  
【대리인코드】 9-1998-000616-9  
【포괄위임등록번호】 1999-043458-0

## 【발명자】

【성명의 국문표기】 유재신  
【성명의 영문표기】 YU, Jae Shin  
【주민등록번호】 770313-2465126  
【우편번호】 151-872  
【주소】 서울특별시 관악구 신림4동 484-31호  
【국적】 KR

## 【발명자】

【성명의 국문표기】 전성배  
【성명의 영문표기】 JUN, Sung Bae  
【주민등록번호】 711010-1057913  
【우편번호】 153-034  
【주소】 서울특별시 금천구 시흥4동 804  
【국적】 KR

## 【발명자】

【성명의 국문표기】 윤경로  
【성명의 영문표기】 YOON, Kyoung Ro

1020010012184

출력 일자: 2001/12/13

【주민등록번호】 641104-1037516  
【우편번호】 135-271  
【주소】 서울특별시 강남구 도곡동 도곡 경남아파트 101동 2004호  
【국적】 KR  
【심사청구】 청구  
【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 허용록 (인)  
【수수료】  
    【기본출원료】 20 면 29,000 원  
    【가산출원료】 1 면 1,000 원  
    【우선권주장료】 0 건 0 원  
    【심사청구료】 10 항 429,000 원  
    【합계】 459,000 원  
【첨부서류】 1. 요약서·명세서(도면)\_1통

**【요약서】****【요약】**

본 발명은 내용 기반 멀티미디어 검색과 브라우징(browsing)을 위한 멀티미디어 요약과 브라우징 시스템에서, 특히 비디오의 효율적인 요약과 이를 이용한 효율적인 검색, 필터링이 가능하도록 텍스트 키프레임을 추출하는 방법에 관한 것이다.

본 발명은 비디오 스트림 요약과 브라우징을 위하여 합성 비디오 텍스트 키프레임을 이용할 수 있도록 합성 텍스트 키프레임을 추출하는 방법에 관한 것으로서, 비디오 스트림에 존재하는 텍스트 영역을 검출하고, 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 검출된 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 비례하여 할당하고, 텍스트 합성 키 프레임을 디스플레이 하는 디바이스나 브라우저의 크기에 알맞게 가중치가 높은 순으로 합성할 텍스트의 개수를 정하거나 합성할 텍스트의 사이즈를 정하고, 이 영역들을 하나의 이미지에 합성하여 일정 비디오 구간을 텍스트 합성 키 프레임으로 대표하는 것을 특징으로 하는 텍스트 합성 키 프레임 추출 방법에 관한 것이다.

**【대표도】**

도 6

**【색인어】**

멀티미디어, 브라우저, 텍스트, 키프레임

【명세서】

【발명의 명칭】

비디오 텍스트 합성 키 프레임 추출방법{METHOD FOR DETECTING CAPTION  
SYNTHETIC KEY FRAME IN VIDEO STREAM}

【도면의 간단한 설명】

도1은 비디오 스트림의 구조적 정보의 예를 나타낸 도면

도2는 본 발명을 적용하는 합성 키프레임을 이용한 비디오 브라우징 인터페이스에 따른 도면

도3은 본 발명의 합성 키프레임 생성의 개념을 설명하기 위한 도면

도4는 본 발명에서 텍스트 추출방법을 설명하기 위한 도면

도5는 본 발명에서 텍스트 합성 키프레임의 개념을 설명하기 위한 도면

도6은 본 발명에 따른 텍스트 합성 키프레임 추출방법의 예를 나타낸 도면

도7은 본 발명에 따른 텍스트 합성 키프레임 추출방법의 다른 예를 나타낸  
도면

도8은 본 발명에서 추출된 영역 내의 글자 크기를 예측하는 방법의 예를 나타낸 도면

**【발명의 상세한 설명】****【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

- <9> 본 발명은 내용 기반 멀티미디어 검색과 브라우징(browsing)을 위한 멀티미디어 요약과 브라우징 시스템에서, 특히 비디오의 효율적인 요약과 이를 이용한 효율적인 검색, 필터링이 가능하도록 키프레임을 추출하는 방법에 관한 것이다.
- <10> 본 발명은 비디오에서 텍스트가 검출된 비디오 스트림에 대해, 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 추출된 텍스트 영역에 할당하여, 상기 가중치를 기반으로 하여 상기 텍스트 중에서 합성할 텍스트를 선정하고, 상기 선정된 텍스트들을 하나의 이미지에 합성하여 일정 비디오 구간을 대표하는 텍스트 합성 키프레임으로 설정함으로써 보다 효율적인 비디오 요약과 검색 및 필터링이 가능하도록 한 비디오 텍스트 합성 키 프레임 추출방법에 관한 것이다.
- <11> 더욱 상세하게는 본 발명은 비디오 스트림 요약과 브라우징을 위하여 합성 비디오 텍스트 키프레임을 이용할 수 있도록 합성 텍스트 키프레임을 추출하는 방법에 관한 것으로서, 비디오에서 텍스트가 검출된 비디오 스트림에 대해, 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 검출된 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 비례하여 할당하고, 텍스트 합성 키 프레임을 디스플레이 하는 디바이스나 브라우저의 크기에 알맞게 가중치가 높은 순으로 합성할 텍스트의 개수를 정하거나 합성할 텍스트의 사이즈를 정하고, 이 영역들을 하나의 이미지에 합성하여 일정

비디오 구간을 텍스트 합성 키 프레임으로 대표하는 것을 특징으로 하는 텍스트 합성 키 프레임 추출 방법에 관한 것이다.

<12> 디지털 비디오 기술의 발달과 이미지/비디오/오디오 인식 기술의 발달로 인하여 사용자들은 원하는 비디오를 원하는 시점에 원하는 부분만을 검색 (searching/filtering)하고 브라우징(browsing) 할 수 있게 되었다.

<13> 비선형적인 비디오 콘텐츠 브라우징(non-linear video content browsing)과 검색을 위하여 가장 기본이 되는 기술은 샷 분할 (shot segmentation) 기법과 샷 클러스터링(shot clustering) 기법이며, 이 두 가지 기술은 멀티미디어 콘텐츠를 구조적으로 분석하는데 있어서 가장 핵심이 되는 기술이다.

<14> 도1에 비디오 스트림의 구조적 정보의 예를 나타내었다.

<15> 시간적인 연속성을 가진 비디오 스트림(Video Stream)에는 도1과 같은 구조적 정보(Structural information)가 존재한다. 일반적으로 비디오 스트림은 장르(genre)에 구분 없이 계층구조(hierarchical structure)를 가지고 있다. 즉, 비디오 스트림은 논리적인 단위인 몇 개의 씬(Scene)으로 나뉘어 지고 각각의 씬은 다수의 서브 씬(Sub-Scene) 또는 샷(Shot)들로 구성되며, 서브 씬은 씬이므로 씬의 속성을 그대로 가진다. 비디오 스트림에 있어서 샷이란 방해 (interruption) 없이 하나의 카메라로부터 얻어진 비디오 프레임들의 시퀀스(sequence of frames)를 의미한다. 대부분의 멀티미디어 인덱싱 시스템들은 샷을 추출하고 추출된 샷을 토대로 하고 기타의 정보를 이용하여 논리적인 단위인 씬들을 검출함으로써 멀티미디어 스트림의 구조적 정보를 인덱싱하게 된다.



<16> 앞에서 설명한 바와같이 샷은 비디오를 분석(analysis)하거나 구성(construction)하는 가장 기본이 되는 단위이다. 또한 일반적으로 비디오에 존재하는 의미적인 구성 요소인 씬(scene)은 이야기 전개 또는 비디오 구성에 있어서 의미적인 구분 요소이며 일반적으로 하나의 씬 내에는 여러 개의 샷이 존재한다. 기존의 비디오 인덱싱 기술은 비디오를 구조적으로 분석하여 샷과 씬을 검출하고 이를 바탕으로 하여 단위 세그먼트인 샷 또는 씬을 대표할 수 있는 키프레임(key frame)들을 추출하여 각각의 샷이나 씬을 대표하고 비디오 요약에 위한 자료로 활용하거나 원하는 위치로 이동하기 위한 수단으로 사용한다.

<17> 이와 같이, 효율적인 비디오 검색과 브라우징을 위하여 비디오 스트림에서 의미적 정보를 표현하는 주요 텍스트 영역이나 뉴스 아이콘 또는 사람의 얼굴 영역 등을 추출하기 위한 연구가 진행 중이며, 이러한 키 영역을 합성하여 새로운 키 프레임을 추출하는 방법이 소개되었다. 합성 키 프레임은 논리적 혹은 물리적 단위의 비디오 스트림의 내용을 씬이나 샷 단위에서 뽑은 키 프레임을 이용하여 합성하는 기술이다. 합성 키 프레임을 이용함으로써 적은 디스플레이 공간에 많은 정보를 표현할 수 있고, 사용자는 콘텐츠의 특정 부분의 내용을 빠르고 쉽게 이해할 수 있고 관심 있는 부분만 선택하여 시청할 수 있다.

<18> 기존의 합성 키 프레임에 관한 기술은 비디오 콘텐츠의 일정 구간을 키 영역이나 키 텍스트를 이용하여 하나의 키 프레임으로 합성한 후, 이 합성 키 프레임을 이용하여 해당 구간을 대표하는 수단으로 이용하는 것에 관한 것이다.

<19> 이 중에서 비디오 텍스트는 일반적으로 콘텐츠 전체 혹은 일부분의 내용을 요약해 주는 특성을 가지고 있기 때문에 사용자에게 콘텐츠에 대한 요약 정보를

제공하는데 있어 매우 중요한 수단이 된다. 기존에는 텍스트 합성 키 프레임의 생성방법에 대한 구체적인 제안이 없었다. 추출된 각각의 텍스트를 중요도에 대한 고려 없이 무순으로 혹은 임의적으로 텍스트 합성 키 프레임을 추출하고 이를 콘텐츠 요약에 이용하면 실제로 중요한 정보를 놓칠 수 있다. 따라서 어떤 텍스트 영역이 실제로 중요한 텍스트 영역인지를 판단하고 이를 어떠한 방법으로 합성하는 것인지는 제한된 공간에서 많은 정보를 전달하기 위한 텍스트 합성 키 프레임을 생성하는 것에 있어서 매우 중요하다.

【발명이 이루고자 하는 기술적 과제】

<20> 본 발명에서는 비디오 콘텐츠에 존재하는 각각의 비디오 텍스트를 자동으로 검출하고, 검출된 텍스트의 각 영역에 대하여 중요도(Weight)를 할당하여 이를 기반으로 중요도가 높은 몇몇 텍스트 영역을 선택하고 이를 하나의 키 프레임으로 합성하는 텍스트 합성 키 프레임 추출 방법을 제안한다. 본 발명에 따른 텍스트 합성 키 프레임은 비디오 콘텐츠의 요약이나 비 선형적인 브라우징의 용도로 사용할 수 있다.

<21> 본 발명에서는 기존의 합성 키 프레임에서 제안되지 않은 텍스트 합성 키 프레임의 생성 방법을 제안한다.

<22> 비디오 텍스트는 비디오 콘텐츠 전 구간에 존재하는데, 기존의 합성 키 프레임에서는 비디오 콘텐츠에서 추출된 비디오 텍스트에 대한 구체적인 합성 방법이 제안되지 않았다. 즉, 추출된 텍스트에 대하여 중요도 판단 없이 텍스트를 합성하게 되어 중요한 텍스트가 합성 키 프레임에 빠질 수가 있다.

<23> 이에 본 발명에서는 추출된 비디오 텍스트에 대하여 각각 중요도를 할당하고 이를 기반으로 중요도가 높은 몇몇 텍스트들을 합성하여 하나의 키 프레임으로 생성하는 텍스트 합성 키 프레임 추출 방법을 제안한다. 제안된 텍스트 합성 키 프레임 추출 방법은 텍스트의 중요도를 판단하여 중요한 텍스트 순으로 텍스트를 합성하기 때문에 비디오 콘텐츠를 좀더 명확하게 요약할 수 있고 사용자의 이해를 높일 수 있는 장점이 있다.

**【발명의 구성 및 작용】**

<24> 본 발명은 비디오에서 텍스트가 추출된 비디오 스트림에 대해, 상기 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 텍스트 요소에 할당하는 단계, 상기 가중치를 기반으로 하여 상기 텍스트 중에서 합성할 텍스트를 선정하는 단계, 상기 선정된 텍스트들을 하나의 이미지에 합성하여 일정 비디오 구간을 대표하는 텍스트 합성 키프레임으로 설정하는 단계; 를 포함하여 이루어지는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법이다.

<25> 또한 본 발명은 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 검출된 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 비례하여 할당하는 단계, 텍스트 합성 키 프레임을 디스플레이 하는 디바이스나 브라우저의 크기에 알맞게 가중치가 높은 순으로 합성할 텍스트의 개수를 정하거나 합성할 텍스트의 사이즈를 정하는 단계, 상기 설정된 텍스트 영역들을 하나의 이미지에 합성하여 일정 비디오 구간을 텍스트 합성 키 프레임으로 대표하는 단계로 이루어진 것을 특징으로 하는 텍스트 합성 키 프레임 추출 방법이다.

<26> 비디오 텍스트의 합성 키 프레임을 이용한 어플리케이션은 비디오 스트림의 특정 범위의 요약이나 비디오 검색을 위한 브라우징 인터페이스를 가진 모든 시스템에서 동작할 수 있다. 도2는 합성 키프레임을 이용한 비디오 브라우징 인터페이스의 예를 보여준다. 도2에서 볼 수 있듯이 여러 개의 키 프레임과 합성 키 프레임을 이용하면 제한된 영역 내에서 많은 정보를 표시할 수 있으므로 비디오 요약에 있어서 유용하며, 비디오 콘텐츠에서 원하지 않는 부분을 스킵하고 관심 있는 부분만을 브라우징하기 위한 툴(Tool)로써 이용이 가능하다

<27> 도3은 합성 키 프레임 생성의 개념을 보여준다.

<28> 대부분의 비디오 인덱싱 시스템에서는 추출된 비디오 스트림의 구조적 구성 요소인 씬과 샷을 대표하기 위하여 키프레임(key frame)을 추출하고, 이를 검색이나 브라우징을 위한 용도로 사용한다. 이를 좀더 효율적으로 하기 위해서 합성 키 프레임 추출 방법이 제시되었다. 합성 키 프레임 추출 방법은 비디오에서 논리적 단위인 씬 혹은 물리적 단위인 샷에서 키 프레임을 추출하고 추출된 키 프레임을 이용하여 논리적 또는 물리적으로 합성하여 사용자에게 제공함으로써 쉽게 비디오 콘텐츠를 이해할 수 있고, 원하는 위치로 빠르게 접근할 수 있게 한다.

<29> 한편, 효율적인 비디오 검색과 브라우징을 위하여 비디오 스트림에서 의미적 정보를 표현하는 주요 텍스트 영역을 추출할 수 있다. 이 기술에 따르면 비디오에 나오는 텍스트의 최소 사각형(MBR) 영역만 추출하여 사용자에게 비디오의 내용을 쉽게 파악할 수 있고, 인덱스 할 수 있는 기능을 제공한다. 또한 유연성

있는 정보 검색과 인덱싱된 정보를 이용하여 네트워크 상에서 원거리 정보 검색도 가능하다.

<30> 도4는 비디오 영상에서 텍스트를 추출하는 방법의 예를 도식화 한 것이다.

<31> 텍스트 추출에는 가로, 세로 에지 히스토그램이 집중적으로 나타나는 성질과 에지 히스토그램이 글자 간격에 따라 반복적으로 커졌다 작아졌다 하는 정보를 이용하여 후보 영역을 추출하고, 후보 영역 중에서 가로 세로 비율이 텍스트임을 만족하며 모션이 적고, 배경과 명도의 차가 큰 칼라가 사용된 영역을 텍스트로 추출하는 방법을 이용한다.

<32> 비디오 콘텐츠에서 나타나는 텍스트가 중요한 텍스트인지 아닌지의 판단은 제한된 영역 내에서 비디오 콘텐츠를 요약하는 비디오 텍스트 합성 키 프레임 생성에서 매우 중요하다.

<33> 따라서, 본 발명에서는 비디오 콘텐츠에서 추출된 텍스트의 각 영역에 대해 추출된 텍스트 영역의 크기, 추출된 영역내의 텍스트의 평균 크기, 디스플레이 존속 시간 등의 정보를 이용하여 중요도를 할당하고, 중요도가 높은 텍스트 영역 몇몇을 선택하여 하나의 합성 키 프레임을 생성함으로써 적은 양의 데이터로 많은 정보를 제공하여 비디오 스트림을 요약하고, 비 선형적인 비디오 브라우징의 수단으로도 이용될 수 있도록 한다.

<34> 본 발명과 관련된 어플리케이션으로는 UMA 응용 어플리케이션이 있다.

<35> 사용자의 단말장치나 사용자 단말장치와 서버를 연결하는 네트워크 환경 등에 따라 사용자가 이용할 수 있는 데이터에 한계가 있을 수 있다. 즉, 어떤 디바

이스를 사용하는가에 따라, 정지영상은 지원되나 동영상 디스플레이는 지원되지 않을 수 있으며, 오디오는 지원되나 영상은 지원되지 않을 수도 있다. 또한 네트워크 연결 방식이나 매체에 따라 전송용량이 부족하여 정해진 시간 내에 전송 받을 수 있는 데이터의 양이 제한될 수 있다. 이러한 다양한 사용자의 환경의 변화에 적응하여, 어떤 멀티미디어 정보를 사용자의 환경에 가장 알맞은 형태로 가공하여 사용자의 편의를 도모하고 정보의 전달 능력을 향상시키고자 하는 모든 어플리케이션들을 Universal Multimedia Access Applications (UMA Applications) 라고 한다.

<36> 예를 들어 디바이스/네트워크 등의 제약조건에 의해 비디오 스트림을 받아 디스플레이 할 수 없는 사용자에게 사용자의 환경이 허락하는 한도내의 축소된 크기와 개수의 텍스트 키 프레임으로 비디오 스트림을 변환시켜 전송함으로써 최소한의 이해를 도모할 수 있다.

<37> 본 발명의 텍스트 합성 키 프레임은 UMA 어플리케이션에 적용되어 전송될 키 프레임을 개수를 줄여서 전송될 데이터의 양은 줄이면서 더 많은 의미적 정보를 제공하는 수단으로 이용될 수 있다.

<38> 본 발명과 관련된 또 다른 어플리케이션으로는 비선형적 비디오 브라우징 어플리케이션이 있다(도2 참조). 전체 비디오 스트림은 요약물 하지 않으면 사용자가 그 비디오 스트림을 이해하기 위하여 그 스트림을 모두 봐야 하는 단점이 있다. 또 원하는 위치로 이동하려고 해도 비디오 스트림내의 그 위치까지 사용자가 직접 탐색해야 하기 때문에 많은 시간이 필요하다. 비디오 스트림을 빠른 시간 내에 검색하고 접근하기 위해서는 비선형적 비디오 브라우징이 이용될 수 있

다. 전체 비디오 내용을 키 프레임을 추출하여 특정 범위 단위로 요약하여 보여주고 그 원하는 위치부터 검색도 가능하다.

<39> 도2에서 살펴보면, 뉴스 비디오의 경우 특정한 기사에 대해서 그 기사 내용을 함축적으로 의미 및 대표하는 키프레임 또는 키 영역이 디스플레이되고 있고, 또한 해당 기사 내용을 텍스트를 기반으로 해서 합성한 텍스트 합성 키프레임이 디스플레이되고 있다. 여기서 텍스트 합성 키프레임을 보면 비디오 디스플레이 뷰에서 앵커가 설명하고 있는 상세 뉴스(해당 기사 구간)에 대하여 그 내용 전체를 요약해서 텍스트로 보여주고 있다.

<40> 이와같은 텍스트 합성 키프레임을 중요도를 기반으로 추출하는 방법을 본 발명에서는 제안한다.

<41> 도5는 본 발명의 텍스트 합성 키 프레임에 대한 개념을 요약한 도면이다. 텍스트 합성 키 프레임은 비디오에서 텍스트 영역이 추출된 비디오 스트림에서 각 영역의 중요도를 계산하여 가중치를 주고, 일정 가중치 이상의 키 영역만으로 새로운 키 프레임을 생성하는 것이 주요 특징이다.

<42> 텍스트 합성 키 프레임은, 추출된 텍스트 영역의 크기(A), 추출된 텍스트 영역 내의 텍스트 요소의 평균 크기(B), 텍스트의 디스플레이 존속 시간(C)를 고려해서 다음과 같이 그 중요도(I)를 얻을 수 있다.

<43> 즉,  $I = A * a + B * b + C * c$       ( $a + b + c = 1$ )

<44> 여기서, a,b,c는 각 요소(A,B,C)에 대한 가중치이며, 가중치의 총합은 '1'이 된다.

- <45>      상기한 바와 같이 가중치를 주는 방법은 추출된 텍스트 영역의 최소 사각형의 크기(MBR)가 크고 추출된 영역 내의 텍스트 요소의 평균 크기가 크고, 텍스트의 디스플레이 존속 시간이 길면 가중치를 크게 주는 방법이다.
- <46>      추출된 텍스트 영역 내의 텍스트 요소의 평균 크기는 도8에서와 같이 히스토그램의 밀도와 크기로 알 수 있다. 즉, 글자 크기가 작으면 각 라인 사이에 가로 에지 히스토그램의 크기가 크게 작아지는 것을 볼 수 있다. 또 세로 에지 히스토그램도 각 글자 사이에서 히스토그램의 크기가 크게 작아진다. 반대로 추출된 영역 내에 글자 크기가 크면 도8에서 볼 수 있듯이 가로 에지 히스토그램의 분포가 넓게 분포되어 있고 중간에 히스토그램의 크기가 갑자기 줄어드는 현상은 발생하지 않는다. 이와 같은 히스토그램의 밀도와 크기의 정보를 이용해 추출된 텍스트 영역내의 텍스트의 평균 크기를 구할 수 있다.
- <47>      추출된 텍스트의 존속 시간은 이전에 추출된 영역과 현재 추출된 영역을 비교하여 구할 수 있다. 추출된 영역의 크기와 위치정보가 유사하고, 에지 히스토그램의 값의 차이가 일정 임계치(threshold) 보다 작으면 두 영역을 같은 텍스트로 판단하여 추출된 텍스트의 디스플레이 존속 시간을 늘인다.
- <48>      텍스트 합성 키 프레임은 비디오 스트림을 구성하고 있는 논리적 단위 '씬', 물리적 단위 '샷' 단위로 생성할 수 있다. 씬과 샷은 도1에서 나타나 있듯이 샷 세그멘테이션 기술로 구분할 수 있다.
- <49>      합성할 비디오 텍스트는 디스플레이 할 디바이스의 크기나 브라우저에서 텍스트 합성 키 프레임의 영역에 따라 합성할 키 프레임의 개수나 각각의 사이즈가 달라질 수 있다. 즉, 디스플레이 할 영역이 커지면 합성할 텍스트의 개수를 늘



리거나 각각의 텍스트의 사이즈를 늘릴 수 있고, 디스플레이 할 영역이 작아지면 합성할 키 프레임의 개수를 줄이거나 각각의 텍스트 사이즈를 줄일 수 있다. 합성할 키 프레임의 개수의 변경이나 각각의 텍스트 영역의 사이즈 변경이 있을 경우 사용자의 가독성(readability)을 고려하여야 한다.

<50> 도6 및 도7은 본 발명에 따른 텍스트 합성 키 프레임 추출의 예를 나타낸 도면으로서, 도6의 경우는 뉴스 비디오에서 특정 기사 구간에 대해서 텍스트 합성 키프레임을 추출하는 방법을 보여주고 있고, 도7의 경우는 쇼 프로그램에서 텍스트 합성 키프레임을 추출하는 방법을 보여주고 있다.

<51> 도6 및 도7에서 보는 것과 같이 특정 범위 내의 텍스트에 대해 각각 중요도를 계산하고, 디스플레이 할 영역의 크기를 고려하여 중요도가 높은 순으로 텍스트를 합성한다.

<52> 도6의 경우를 본다면, 뉴스 비디오 콘텐츠에서 특정 기사에 해당하는 구간 내의 텍스트 영역을 모두 추출한 다음, 검출된 텍스트 영역에 대한 중요도 가중치를 그 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 비례하여 할당하고, 브라우저(디스플레이) 크기에 맞게 가중치가 높은 순서로 합성할 텍스트의 갯수(또는 사이즈)를 정해서 이 정해진 갯수(또는 사이즈)의 텍스트 영역들을 하나의 이미지에 합성하여 텍스트 합성 키프레임으로 그 기사의 내용을 함축적으로 보여주는 것이다.

<53> 도7의 경우는 쇼 비디오 콘텐츠에서 특정 구간에 대하여 텍스트 영역을 추출한 다음, 앞에서 설명한 뉴스 비디오 콘텐츠처럼 중요도 가중치와 브라우저의

크기 등을 고려해서 적절한 갯수(또는 사이즈)의 텍스트 영역들을 하나의 이미지로 합성해 넣으로써, 텍스트 합성 키 프레임을 추출한 것을 보여준다.

**【발명의 효과】**

<54>        본 발명에서는 추출된 비디오 텍스트에 대하여 각각 중요도를 할당하고 이를 기반으로 중요도가 높은 몇몇 텍스트들을 합성하여 하나의 키 프레임으로 생성하는 텍스트 합성 키 프레임 추출 방법을 제안한다. 제안된 텍스트 합성 키 프레임 추출 방법은 텍스트의 중요도를 판단하여 중요한 텍스트 순으로 텍스트를 합성하기 때문에 비디오 콘텐츠를 좀더 명확하게 요약할 수 있고 사용자의 이해를 높일 수 있는 장점이 있다.

/ 1020010012184

출력 일자: 2001/12/13

**【특허청구범위】****【청구항 1】**

비디오에서 텍스트가 검출된 비디오 스트림에 대해, 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 텍스트 요소에 할당하는 단계, 상기 가중치를 기반으로 하여 상기 텍스트 중에서 합성할 텍스트를 선정하는 단계, 상기 선정된 텍스트들을 하나의 이미지에 합성하여 일정 비디오 구간을 대표하는 텍스트 합성 키프레임으로 설정하는 단계; 를 포함하여 이루어지는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법.

**【청구항 2】**

제 1 항에 있어서, 상기 중요도 가중치는 텍스트 요소로서 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 대하여 할당하는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법.

**【청구항 3】**

비디오에서 텍스트가 검출된 비디오 스트림에 대해, 상기 검출된 각각의 텍스트 영역에 대한 중요도 가중치를 검출된 텍스트 영역의 크기, 텍스트 영역에서의 평균 텍스트 요소의 크기, 텍스트 영역의 디스플레이 지속시간에 비례하여 할당하는 단계, 텍스트 합성 키 프레임을 디스플레이 하는 디바이스나 브라우저의 크기에 알맞게 가중치가 높은 순으로 합성할 텍스트의 개수나 사이즈를 정하는 단계, 상기 설정된 텍스트 영역들을 하나의 이미지에 합성하여 일정 비디오 구간

을 텍스트 합성 키 프레임으로 대표하는 단계를 포함하여 이루어진 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법.

**【청구항 4】**

제 3 항에 있어서, 상기 텍스트 영역의 크기는 비디오 텍스트 영역을 감싸는 최소사각형(MBR)의 크기로 판단함을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출 방법.

**【청구항 5】**

제 3 항에 있어서, 상기 텍스트 영역에서의 평균 텍스트 요소의 크기는 추출된 텍스트 영역 내에서 에지(Edge)의 가로, 세로 히스토그램의 밀도와 크기 등의 정보를 가지고 영역내의 평균 텍스트의 크기를 판단 함을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출 방법.

**【청구항 6】**

제 3 항에 있어서, 상기 텍스트 영역의 디스플레이 지속시간은 이전에 검출된 텍스트 영역과 현재 검출된 텍스트 영역의 유사성을 판단하여 같은 텍스트라고 판단되면 디스플레이 지속시간을 늘림을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출 방법.

**【청구항 7】**

제 3 항에 있어서, 상기 일정 가중치 이상의 텍스트 영역을 선택하는 방법은 검출된 텍스트 영역의 크기나, 추출된 텍스트 영역 내의 텍스트의 평균 크기

나, 추출된 텍스트 영역의 디스플레이 시간에 비례하여 가중치를 높이는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출 방법.

**【청구항 8】**

제 1 항에 있어서, 상기 텍스트 합성 키 프레임으로 요약될 일정 구간은 비디오의 논리적 편집 단위인 '씬'이나 비디오의 물리적 편집 단위인 '샷' 인 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출 방법.

**【청구항 9】**

제 1 항에 있어서, 상기 합성할 텍스트의 개수는 텍스트 합성 키 프레임을 디스플레이하는 디바이스나 브라우저의 크기에 비례하여 늘리거나 줄이는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법.

**【청구항 10】**

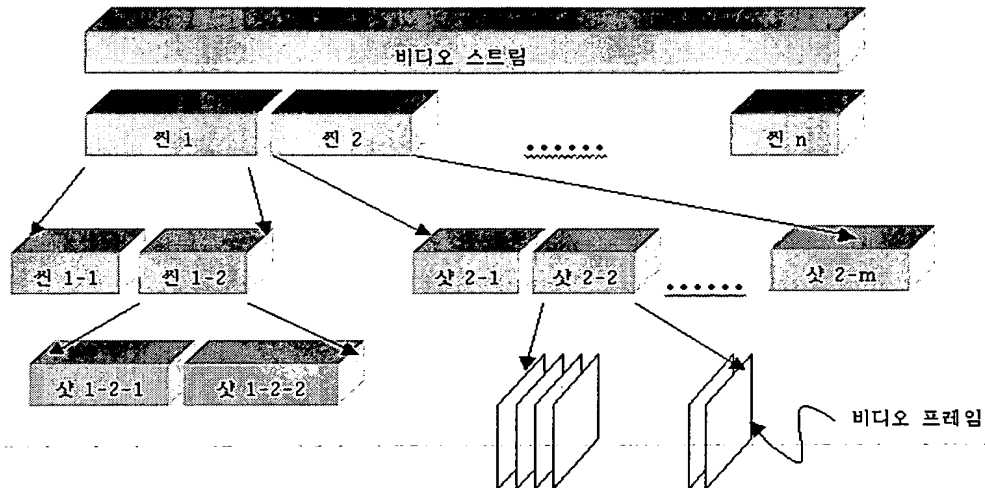
제 1 항에 있어서, 상기 합성할 텍스트의 사이즈는 텍스트 합성 키프레임을 디스플레이하는 디바이스나 브라우저의 크기에 비례하여 늘리거나 줄이는 것을 특징으로 하는 비디오 텍스트 합성 키 프레임 추출방법.

1020010012184

출력 일자: 2001/12/13

## 【도면】

【도 1】



【도 2】

비디오 디스플레이 뷰

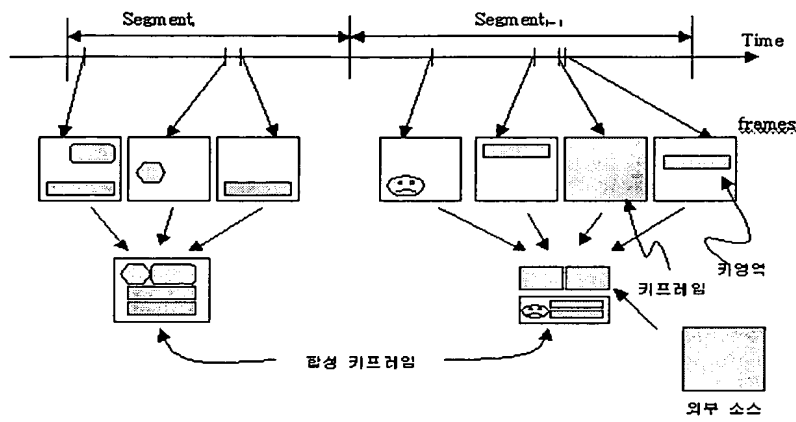
텍스트 합성 키프레임 뷰



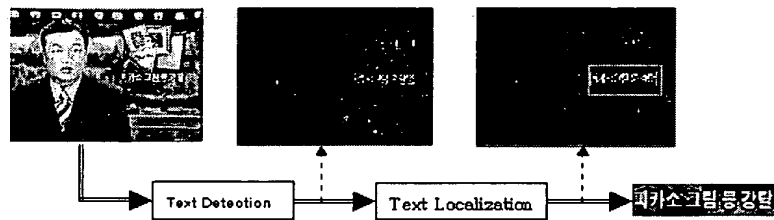
키프레임/키 영역 뷰



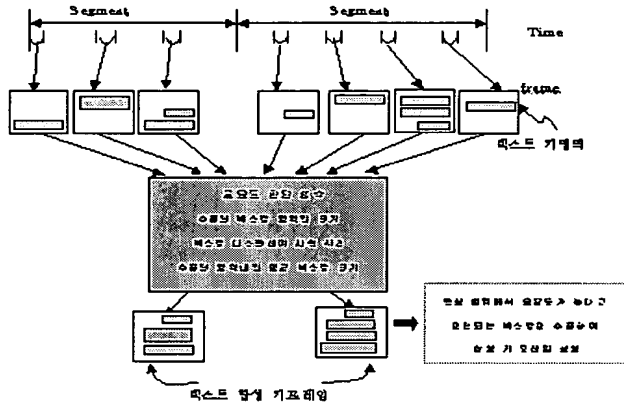
【도 3】



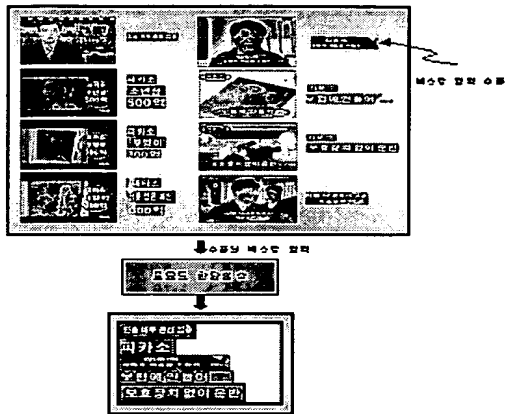
【도 4】



【도 5】



【도 6】



【도 7】



【도 8】

